REMARKS

We appreciate the Examiner taking the time to speak with us on May 14, 2008. Applicants have had an opportunity to carefully consider the Examiner's comments set forth in the Interview as well as in the Office Action of January 22, 2008. Per the Examiner's recommendation, Applicants have amended claims 1, 11 and 20 to fully clarify the Jaccard limitation. Claims 1-4, 7-14 and 17-20 remain in this application. Claims 5, 6, 15 and 16 have been canceled.

Reconsideration of the Application is requested in view of the amendments and comments herein.

I.   The Office Action

Claims 1-2, 4-7, 10-12, 14-17 and 20 are rejected under 35 U.S.C. 103(a) as being unpatentable over Steven J. Simske (U.S.PG Pub. No. 2004/0133560) in view of Taher et al. (NPL "Evaluating Strategies for Similarity Search on the Web" ACM May 7-11, 2002, PP 1-23), further in view of Henkin et al. (U.S. PG Pub No. 2002/0107735).

Claims 3 and 13 are rejected under 35 U.S.C. 103(a) as being unpatentable over Steven J. Simske in view of Taher et al., further in view of Henkin et al. as applied to claims 1-2, 4-7, 10-12, 14-17 and 20 above, further in view of Rie Kubota (U.S. Patent No. 6,041,323).

Claims 9 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Steven J. Simske in view of Taher et al., further in view of Henkin et al. as applied to claims 1-2, 4-7, 10-12, 14-17 and 20 above, further in view of Drissi et al. (U.S. PG Pub No. 20003/0149686).

Claims 8 and 18 are rejected under 35 U.S.C. 103(a) as being unpatentable over Steven J. Simske.

II.   Rejection of Claims 1-2, 4-7, 10-12, 14-17 and 20 under 35 U.S.C. 103(a)

The Examiner has rejected claims 1-2, 4-7, 10-12, 14-17, and 20 under 35 U.S.C. 103(a) as being unpatentable over Simske (U.S. PG Pub No. 2004/0133560) in view of Taher et al. (NPL "Evaluating Strategies for Similarity Search on the Web" ACM, May 7-11, 2002, PP 1-23) further in view of Henkin et al. (U.S. PG Pub No. 2002/0107735). This rejection should be withdrawn for at least the following reasons.

Simske, Taher, and Henkin, individually or in combination, do not teach or suggest the subject invention as set forth in the subject claims.

As amended, claim 1 (and similarly claim 11) recite that if a first computed percentage does not indicate that that the first document is included in a second document, a third percentage is computed using the Jaccard distance measure. If the Jaccard similarity distance measure is greater than about 90 percent, the second document is likely a revision of the first document, and if the Jaccard similarity distance measure is less than about 90 percent, said measure represents the similarity between said first and second document. If the third computed percentage indicates that the first document is a revision of the second document, a fourth percentage is computed indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list. Simske, Taher and Henkin individually or in combination do not teach or suggest such claimed aspects of the subject invention.

In particular, the Examiner admits Simske does not teach "if the first computed percentage does not indicate that the first document is included in the second document, computing a third percentage using the Jaccard distance measure." However, the Examiner argues that such a feature is found in Taher. Applicant respectfully disagrees with the Examiner. Although Taher does teach of a Jaccard coefficient for measuring the similarity of document bags, Taher does not teach or suggest that the Jaccard coefficient is to be used as a third computed percentage only if the first computed percentage does not indicate that the first document is included in the second document. Thus, the Jaccard coefficient is not utilized in the same way as recited in the subject claims.

Taher employs the Jaccard coefficient generally as a metric for measuring the similarity of document bags. If the Jaccard coefficient was to be used in the same way in the present application, it would change the entire method, since a main feature of the present application is that document similarity is determined using lists of rated keywords. As recited in the subject claims, the Jaccard similarity distance measure is only implemented if the first computed percentage does not indicate that the first document is included in the second document to determine if the second document is a

revision of the first document. The combination of Taher and Simske does not provide or fairly suggest such a feature.

The Examiner supports the combination by stating that the Taher teachings would have allowed Simske to provide reduced costs in both time and resources and providing efficient and quality results. Applicant asserts that the Examiner's reason for combining Taher and Simske is invalid. The Examiner fails to provide any rationale as to why or how the combination of Taher and Simske would reduce costs and provide efficient and quality results. Moreover, there is no suggestion to combine the teachings and suggestions of Taher and Simske, except from using Applicant's invention as a template through a hindsight reconstruction of Applicant's claims. Rejections on obviousness grounds cannot be sustained by mere conclusory statements; instead, there must be some articulated reasoning with some rationale underpinning to support the legal conclusion of obviousness. The Examiner's statement, that it would have been obvious to combine the Simske and Taher reference because it would increase efficiency and results while reducing cost, is an example of such a conclusory statement. There is no teaching in Taher that would support such a conclusion.

The Examiner further states that neither Simske nor Taher disclose "if the third computed percentage indicates that the first document is a revision of the second document, computing a fourth percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list." However, the Examiner states such a feature is found in Henkin, specifically in paragraphs 0229 and 0288. Applicant respectfully disagrees. Paragraph 0229 of Henkin discloses a match type portion where a user can specify what type of match (e.g. exact or fuzzy) is required. If a fuzzy match is to be in effect, a threshold value may be used for determining the minimum percent threshold value required for a "fuzzy" match.

Paragraph 0288 describes calculating the percentage of matched words in a retrieved text portion if no negative words are found within a set number of words from a currently selected word. Applicant is unclear how the Examiner is comparing such a feature with that of the present invention. Nowhere does Henkin disclose or suggest a computed third percentage capable of indicating if the first document is a revision.

Henkin uses negative words and matched words to compare documents, not rated keywords as does the present invention. Therefore, it is not possible for Henkin to compute a fourth percentage indicating what percentage of keyword ratings along with a set of their neighboring keyword ratings in the second list also exist in the first list.

Moreover, Simske does not teach verifying the inclusion of the first document in the one or more disparate documents by computing a second percentage for each of the one or more disparate documents. Nor does Simske teach using the second computed percentage to indicate what percentage of keyword ratings, along with a set of their neighboring keyword ratings in the first list, also exist in the list for at least one of the one or more disparate documents when the first computed percentage indicates that the first document is included in at least one of the one or more disparate documents.

The Examiner refers to step 601 of Simske, which describes a clustering process. Simske describes calculating a "shared word weight" that correlates the two documents and is the sum of all weight values divided by the number of documents to produce a mean value of all relevant word weights. However, Simske only describes the calculation of a single percentage. Nowhere in step 601 of Simske does it disclose calculating a second percentage for each document to determine if the first document is included in the one or more disparate documents, as recited in the subject claims.

With respect to independent claim 20, Applicant refers the Examiner to the above comments in reference to independent claims 1 and 11. Claim 20 discloses an article of manufacture for computing a measure of similarity between a first document and one or more disparate document. Simske, in view of Taher, further in view of Henkin do not individually or in combination teach or suggest the claimed aspects of the subject invention. In addition to the above comments, Simske does not teach or suggest a fourth computed percentage to specify the measure of similarity except when: (i) the fourth computed percentage is greater than the second computed percentage; (ii) the first list of rated keywords is identified using OCR; (iii) the fourth computed percentage is greater than fifty percent; and (iv) less than twenty percent of the keywords in the first list of keywords are in the second list of keywords are in the second list of keywords, as recited in claim 20.

The Examiner states that Simske discloses that if any documents being considered are paper-based, tools such as a zoning analysis engine in combination with an optical character recognition (OCR) engine may be used to convert the paper-based document to an electric document. (Simske [0016]). This disclosure, however, does not teach or suggest a fourth computer percentage to specify the measure of similarity between the first rated keywords and the list of rated keywords for the second document. The fact that Simske identifies its keywords in the documents with OCR works to make the present invention more nonobvious since Simske did not contemplate the need for a fourth computed percentage if a third percentage indicates that the first document is a revision of a second document. The Examiner admits that Simske has no need for using a fourth computed percentage, therefore Simske necessarily fails to teach or suggest such a feature. The present application does not disclose that if OCR is used, no fourth percentage is necessary, just that the fourth computed percentage will not specify the measure of similarity if the first list of rated keywords is identified by OCR.

For at least the aforementioned reasons, Simske in view of Taher, further in view of Henkin individually or in combination do not teach or suggest the subject invention as recited in independent claims 1, 11, 20 (or claims 2-4, 7-10, 12-14 and 17-29 which respectively depend therefrom). Therefore, this rejection should be withdrawn.

III. Rejection of Claims 3 and 13 Under 35 U.S.C. 103(a)

The Examiner has rejected claims 3 and 13 under 35 U.S.C. §103(a) as being unpatentable over Simske (U.S. PG Pub No. 2002/0107735) in view of Taher et al. (NPL "Evaluating Strategies for Similarity Search on the Web" ACM, May 7-11, 2002, PP 1-23), further in view of Henkin et al. (U.S. PG Pub No. 2002/0107735) as applied to claims 1-2, 4-7, 10-12, 14-17, and 20 above, further in view of Kubota (U.S. Patent No. 6,041,323). This rejection should be withdrawn for at least the following reasons. Claims 3 and 13 depend from independent claims 1 and 11 respectively, and the combination of references cited by the Examiner and referred to above, do not make up for the aforementioned deficiencies of Simske, Taher, and Henkin regarding the present application. Thus, for at least the reasons discussed above with respect to claims 1 and

11, the combination of Simske, Taher, Henkin and Kubota do not teach or suggest the subject claims. Accordingly, this rejection should be withdrawn.

IV.   Rejection of Claims 9 and 19 Under 35 U.S.C. 103(a)

The Examiner has rejected claims 9 and 19 under 35 U.S.C. 103(a) as being over Simske (U.S. PG Pub No. 2004/0133560) in view of Taher (NPL "Evaluating Strategies for Similarity Search on the Web" ACM, May 7-11, 2002, PP 1-23), further in view of Henkin et al. (U.S. PG Pub No. 2002/0107735) as applied to claims 1-2, 4-7, 10-12, 14-17, and 20 above, in view of Drissi et al. (U.S. PG Pub No. 20003/0149689). This rejection should be withdrawn for at least the following reasons. Claims 9 and 19 depend from independent claims 1 and 11 respectively. The combination of Simske, Taher and Henkin do not teach or suggest the subject invention as is described above and Drissi does not make up for the deficiencies referred to above. For at least the reasons discussed above, the rejection of claims 9 and 19 should be withdrawn.

V.   Rejection of Claims 8 and 18 Under 35 U.S.C. 103(a)

The Examiner has rejected claims 8 and 18 under 35 U.S.C. 103(a) as being unpatentable over Simske (U.S. PG Pub No. 2004/0133560). This rejection should be withdrawn for at least the following reasons. Claims 8 and 18 depend from independent claims 1 and 11 respectively and, as noted above, are now in condition for allowance. Thus, for at least the reasons discussed above with respect to claims 1 and 11, Simske does not teach or suggest the subject claims. Accordingly, the rejection of claims 8 and 18 should be withdrawn.

CONCLUSION

For the reasons detailed above, it is submitted all claims remaining in the application (Claims 1-4, 7-14 and 17-20) are now in condition for allowance. The foregoing comments do not require unnecessary additional search or examination.

☒ Remaining Claims, as delineated below:

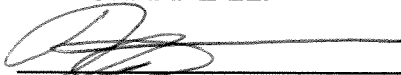| (1) FOR	(2) CLAIMS REMAINING AFTER AMENDMENT LESS HIGHEST NUMBER PREVIOUSLY PAID FOR		(3) NUMBER EXTRA
TOTAL CLAIMS	16	- 20 =	0
INDEPENDENT CLAIMS	3	- 3 =	0

☒ This is an authorization under 37 CFR 1.136(a)(3) to treat any concurrent or future reply, requiring a petition for extension of time, as incorporating a petition for the appropriate extension of time. The fee for the one-month extension of time is being submitted via EFS web.

☒ The Commissioner is hereby authorized to charge any filing or prosecution fees which may be required, under 37 CFR 1.16, 1.17, and 1.21 (but not 1.18), or to credit any overpayment, to Deposit Account 06-0308.

In the event the Examiner considers personal contact advantageous to the disposition of this case, he/she is hereby authorized to call Mark Svat, at Telephone Number (216) 861-5582.

Respectfully submitted,

FAY SHARPE LLP

5/20/08
Date

Mark S. Svat, Reg. No. 34,261
Kevin M. Dunn, Reg. No. 52,842
1100 Superior Avenue, Seventh Floor
Cleveland, OH 44114-2579
216-861-5582

CERTIFICATE OF TRANSMISSION	
I hereby certify that this correspondence (and any item referred to herein as being attached or enclosed) is (are) being ☒ transmitted to the USPTO by electronic transmission via EFS-Web on the date indicated below.	
Express Mail Label No.:	Signature: _Elaine M. Checovich_
Date: 5-20-08	Name: Elaine M. Checovich

N:\XERZ\201374\KAT0000105V001.DOCX